U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE

OFFICE NOTE 306

DATA PROCESSING, AND QUALITY CONTROL
FOR OPTIMUM INTERPOLATION ANALYSES AT THE
NATIONAL METEOROLOGICAL CENTER

GEOFFREY J. DIMEGO AND PATRICIA A. PHOEBUS
DEVELOPMENT DIVISION

AND

JAMES E. MCDONELL
AUTOMATION DIVISION

FEBRUARY 1985

DATA PROCESSING, AND QUALITY CONTROL
FOR OPTIMUM INTERPOLATION ANALYSES AT THE
NATIONAL METEOROLOGICAL CENTER

Geoffrey J. DiMego, Patricia A. Phoebus, James E. McDonell
National Meteorological Center
Washington, DC   USA

1.   INTRODUCTION

In this paper we will describe the procedures for collecting, prepro-

cessing, and quality controlling the data used for meteorological analyses at

the National Meteorological Center (NMC). The discussion is focused on NMC's

global optimum interpolation (GOI) analysis system which is used for both

operational medium-range forecasting and data assimilation. We will not discuss

the data aspects of our Limited-area Fine-Mesh system or the previously opera-

tional HOUGH system, which is still used but only as a backup procedure.

The evolution of OI analyses at NMC has been documented in the series

of articles by Bergman (1979), McPherson et al. (1979), Kistler and Parrish

(1982) and Dey and Morone (1985). However, very little of what is presented

here is contained in those papers. We intend to address some essential but

unglamorous aspects of a numerical weather prediction operation that seldom

appear in the literature. The paper is organized in the order that the tasks

are normally performed. We begin with an overview of the data collection

procedures in Section 2, followed by the discussion of the data preprocessor

for the OI analysis in Section 3. We discuss the details of the generation

and preliminary quality checking of forecast errors in Section 4, and the

pre-analysis buddy check in Section 5. Differences between the GOI and the

new regional optimum interpolation (ROI) system are discussed in Section 6.

2.   DATA COLLECTION AND PRELIMINARY PROCESSING

2.1   Collection

While the primary source of data is the Global Telecommunications System (GTS), there are various other sources of information - regional, governmental, and military, for example - that must be considered. The formidable task of collecting, processing and collating the meteorological data base is performed by NMC's Automation Division. At present, almost all of the communication and collection functions are performed by a pair of IBM 4341 computers.

Each incoming bulletin has its receipt time logged and is then staged to a holding file. Those bulletins of the same data type - for example, surface reports - are chained together such that they constitute a logical file. The receipt time accompanies each report at every stage of additional processing. Headings which are unrecognizable - i.e., those which have no matching entry in the switching directory - are displayed for possible correction and re-entry to the system for processing. In the case of the upper-air reports, the processor which examines that logical chain of bulletins is invoked frequently. This makes the reports available for manual inspection and correction via a KCRT. Additionally, the upper-air processor generates queues of unprocessable reports for visual inspection. Copies of the other logical files of raw data are transferred from the 4341 system to the front-end computer system (NAS9040) at intervals of about 20 minutes.

The processed upper-air file is transferred to the 9040 at the beginning of each analysis suite, and the processors which handle the other types of data are invoked to form the basic data sets which provide the input to the analysis data preprocessor. The locally generated satellite soundings and

cloud-track wind estimates are produced in the 9040 system, and are also available as part of the basic data set. A list of these data sets, their contents and their relevant time periods are given in Table 1.

Table 1. Basic Data Sets for OI Analyses at NMC

Name	Contents	Valid Times (GMT)[1]
ADPSFC.TxxZ	Land station (SYNOP) N. American hourlies	xx = 00, 06, 12, 18
SFCSHP.TxxZ	Buoys (fixed, drifting) Ships (fixed, moving) MARS (marine reporting stations)	xx = 00, 06, 12, 18
SFCBOG.TxxZ	Sea-level pressure bogus Satellite moisture bogus Australian sea-level bogus	xx = 00, 06, 12, 18
ADPUPA.TxxZ	Land station upper air Ship upper-air Reconnaissance data	xx = 00, 06, 12, 18
UPABOG.TxxZ	Height bogus (250 mb)	xx = 00, 12
AIRCFT.TxxZ	Aircraft reports (AIREP) Constant-level balloons	xx = 00, 06, 12, 18
AIRCAR.TxxZ	Aircraft reports (ACARS[2])	xx = 00, 06, 12, 18
SATWND.TxxZ	Satellite wind estimates	xx = 00, 12, 18
TSFLAG(CxxG)	Monitor flags for satellite soundings	xx = 00, 06, 12, 18
TSDCTY(CxxG)	Directory for satellite sounding file	xx = 00, 06, 12, 18
TOVS.NMCEDS	Temperature soundings from polar-orbiting satellites	

[1]Valid observation times within main synoptic data set

xx = 00	2100-0259 GMT	xx = 12	0900-1459 GMT
xx = 06	0300-0859 GMT	xx = 18	1500-2059 GMT

[2]Arinc Communications Addressing and Reporting System

2.2   Formatting

With the exception of the TSFLAG, TSDCTY and TOVS (TIROS Operational Vertical Sounding) files, all of the basic data sets in Table 1 are packed in a uniform format designed for maximum efficiency of storage. This format is described in NMC Office Notes ON 29 (1969) and ON 124 (1973). Besides the report identification information, there are several categories of data accommodated by the format. The categories and pertinent parameters for ON 29/124 format data are described in Table 2. Within a single report there can be more than one category of information, and there can be any number of levels of information within a given category. In general, however, all of the information concerning an observation at a particular time has been combined into a single report.

Table 2.   Data Categories and Contents for NMC ON 29/124 Format

Category	Description	Contents
1	Mandatory Level	Z, T, Tdd, DD, FF, QM
*2	Significant Level	P, T, Tdd, QM
*3	Winds-by-Pressure	P, DD, FF, QM
*4	Winds-by-Height	Z, DD, FF, QM
5	Tropopause Level	P, T, Tdd, DD, FF, QM
6	Flight-Level Winds	Za, T, Tdd, DD, FF, QM
8	Miscellaneous	Bogus data, cloud-drift wind P
51	Surface data	Po, P*, T, Tdd, DD, FF, QM

* First level reserved for surface

Abbreviations and units

P	pressure (m)	T	temperature (K)
Po	sea-level pressure (mb)	Tdd	dew-point depression (K)
P*	station pressure (mb)	DD	wind direction (degrees)
Za	pressure altitude (m)	FF	wind speed (knots)
Z	geopotential height (m)	QM	quality marks (EBCDIC)

During the processing of the raw-data holding files into ON 29/124 format several data selection steps are performed. Duplicate reports are removed retaining the report with the largest number of data groups present.

For surface land stations, which report more frequently than every six hours, the observation nearest the main synoptic time is chosen. This means that an observationist at an intermediate synoptic-time provides "back-up" for a missing report at one of the main synoptic times. This is also true of surface observations which are available at nonstandard times. The arrangement is such that the Global Data Assimilation System (GDAS) will utilize a report only once in any 6 hour cycle. In addition, the station elevation and location are added to each report at this time from a master station dictionary containing information for each reporting location. For radiosonde reports, a code indicating the instrument type is also included.

2.3 Consistency Checks

The coded quality marks of ON 29/124, listed in Table 3, indicate the results of certain objective consistency checks made on the data as they are processed into the basic data sets. In the following sections, we will discuss the consistency checks and flags which are most relevant to the OI analyses.

2.3.1 Radiosonde Data

Radiosonde sounding data are subjected to the following consistency checks. All mandatory-level data are first checked for reasonable meteorological values. The reported values must fall within the range specified in Table 4; if they do not, they are assigned quality mark "B". Similarly, significant level temperature data are checked with limits found by interpolating between levels.

Table 3. Relevant Quality Marks for ON 29/124 Format Data

A. Universal Quality Marks

blank/$	Not specified
H	Monitor requests retention
P	Monitor requests non-use

B. Upper-Air Parameters Z, T and wind Categories 1 through 5

Code	Meaning
A/I	Passed vertical consistency check with tight limits
B/J	Failed gross-error check and not recomputed
C/K	Parameter was missing and has been recomputed
D/L	Failed vertical consistency check with tight limits, passed with loose limits
F/N	Failed vertical consistency check with loose limits

C. Surface Parameters P, T and Z Categories 2 through 4

		ROI Code
U/2	Surface data from Parts A and B disagree; Part A is chosen	2
V/3	Surface data from Parts A and B agree	1

D. Surface Marine Parameters P* and Wind Category 51

A	Ship or buoy wind measurement by anemometer	1
D	Unreliable Po value from a ship report	9

E. Sea-Level Pressure Parameter Po Category 51

A	Good agreement between Po and P*	1
B	Disagreement between Po and P*	9
C	Missing P* where one is normally available	3
D	Fair agreement between Po and P* (high elevation)	2

Table 4. Limits for Rawinsonde Data Checks

Level	Reference Height	D-Val Meters Low	D-Val Meters High	Temp. °C Low	Temp. °C High	Max Wind Speed Knots	Temp. Diff. °K
1000	113	- 671	488	-65	60	60	1.1
850	1457	- 823	396	-50	45	80	.9
700	3016	- 915	457	-50	30	100	1.5
500	5572	-1067	549	-57	5	150	3.9
400	7181	-1311	610	-66	-10	175	3.2
300	9159	-1433	793	-72	-20	225	5.4
250	10359	-1524	915	-76	-25	225	4.7
200	11784	-1524	915	-78	-30	225	3.9
150	13618	-1524	915	-85	-30	200	3.1
100	16206	-2206	1294	-95	-30	175	3.9
70	18486	-1990	1110	-95	-25	150	4.6
50	20632	-2230	970	-95	-15	150	3.4
30	23893	-2890	1610	-95	- 5	150	4.9
20	26481	-2980	1520	-95	5	150	3.3
10	31053	-4050	1950	-95	15	150	-

Non-mandatory level winds are checked only for gross errors, and are used in the following vertical consistency checks for the mandatory-level winds. Let DDM and FFM be the mandatory-level wind direction and speed, respectively, and let DDS and FFS be the corresponding values for the nearest significant-level wind. Winds-by-height (when available) are used first, and are used for testing at all levels within 3000 meters of the level being tested. Now let FFMEAN = 1/2 (FFM + FFS) be the mean speed, let DIFDD = DDM - DDS be the direction difference, and let FFDIF = FFM - FFS be the speed difference. If any of the following are true, the wind is said to pass the vertical consistency check and is given an "A" quality mark:

(1) FFMEAN < 30 and FFDIF < 50

(2) FFMEAN < 39 and FFDIF < 50 and DIFDD < 70

(3) FFMEAN < 39 and FFDIF < 50 and DIFDD < 55

(4) FFDIF < 50 and DIFDD < 40

All other cases result in a failure of the check and an "F" quality mark.

If winds-by-height are not available, winds-by-pressure are considered. Only wind reports between 600 mb and 125 mb are used to check mandatory level winds between 500 and 150 mb. The mandatory-layer midpoints are used as the demarkation points for determining which mandatory level is to be checked; i.e., winds between 600 and 450 mb are used to check the 500 mb level, winds between 450 and 350 mb are used to check the 400 mb level, and so forth. The same criteria are used as in the wind-by-height tests, except that FFDIF is tested against 80 instead of 50. Untested mandatory-level winds are checked against the next mandatory level above, unless the next mandatory level below has been checked and passed. In either case, the test criteria are those imposed when checking against winds-by-pressure, with the additional requirement that the magnitude of the vector difference be less than 80 as well.

The heights and temperatures are tested in the following manner. The mandatory- and significant-level temperatures are merged and checked for super-adiabatic lapse rates. The temperature at the top of an unstable layer is re-calculated for internal use only. In addition, missing mandatory-level temperatures are computed from bracketing significant-level temperatures and pressures, provided they are within 100 mb of each other. The allowable pressure difference for mandatory levels above 100 mb is 15 mb. Similarly, missing mandatory-level heights are computed by hydrostatic integration - provided the temperature and height are available at the next lower level and the temperature is available at the level in question. Calculated values receive a quality mark of "C".

Mandatory-level heights and temperatures are next checked for vertical consistency. Using values at the base and top of each mandatory layer, two estimates of the mean virtual potential temperature are computed: θ_T from

the temperature and moisture data, and θ_Z from the height data. A layer
is assumed to be vertically consistent if both θ_T and θ_Z increase over
the layer below, and the absolute value of their difference, $|\theta_T - \theta_Z|$,
is less than the value given in Table 4. Such a layer would receive an "A"
quality mark.

The layers are tested upwards from the surface, proceeding until a
violation is encountered. Then, a series of tests is performed to determine
which parameter is most likely in error. All of the tests involve calculating
a trial value and retesting the layer with it. A successful test with the
trial value indicates that the reported value is the source of the problem.
The testing proceeds in the following manner.

If θ_T decreases with height and $|\theta_T - \theta_Z|$ is excessive,
then a trial temperature is calculated from the next lower layer, and the
tests are retried. If a successful test is achieved by using the trial tempera-
ture, the reported temperature is marked as a failure using a quality mark of
"F". Similarly, if θ_Z decreases with height and $|\theta_T - \theta_Z|$ is
excessive, then a trial height is calculated for the lower layer and the
tests are retried. If a successful test results by using the trial height,
the reported height is marked as a failure using a quality mark of "F".

If both θ_T and θ_Z increase with height, and $|\theta_T - \theta_Z|$ for both
layers is excessive, then it is assumed that the error is in the middle tempera-
ture or height. First, a trial temperature is calculated and compared to the
reported temperature. If the difference is greater than 3°C, the tests are
retried with the calculated temperature. If the tests are successful, the
reported temperature is marked a failure "F". If the difference is less than
3°C, or if the tests with the calculated temperature still fail, then a trial

height is calculated and compared to the reported height. If the difference is at least 75 m, the tests are retried using the trial height. If successful, the reported height is given a quality mark of "F". If both trial values agree with the reported values, then the original values are retested using loosened limits on $|\theta_T - \theta_Z|$ (values in Table 4 increased by 25%). If the test is successful, both reported values are given "D" quality marks. If the tests are still unsuccessful, the testing method has failed and the reported height and temperature are flagged as not checked, " " (a blank). They are not flagged as failures since both θ_T and θ_Z increase with height.

In general, recognizing vertically consistent data is fairly simple and useful; these tests accomplish that. Determining which parameter is in error is much less reliable. For example, it is possible to get a trial value that causes the reported value to be marked a failure, when in fact another parameter is incorrect. This usually happens only when there are insufficient significant-level temperatures available for testing. Because of this uncertainty, the procedure of flagging the results of the tests - but leaving the reported values intact - has been adopted.

2.3.2 Aircraft Data

Aircraft wind reports are checked for wind directions in the range 0-360 degrees and wind speeds between 0 and 300 knots. If this check is failed, or if other decoding problems arise, the report is not accepted for processing into ON 29/124 format. Instead, it is written to a special error file which is examined periodically by the monitor. The ACARS program is a U.S. effort where specially equipped domestic aircraft transmit pressure altitude, wind and temperature information at a higher frequency in time during ascent (following takeoff) and descent (for landing) than during normal level flight. The

ACARS reports are treated in the same manner as other aircraft reports. At this time, there are no quality or consistency checks performed on the satellite cloud-track wind estimates.

2.3.3 Surface Data

Except for the "H" and "P" flags, the GOI does not examine quality marks for surface data. In fact, surface land data are not used at this time in the GOI. On the otherhand, the ROI honors all quality marks for surface data as indicated by the last column of Table 3.

The quality marks listed in Table 3E for sea-level pressure represent the results of the following consistency checks performed on surface land stations which report both sea-level pressure and station pressure. Let ELEV be the station (m), P^* be the reported station pressure, T^* be the reported surface temperature, and P_0 be the reported sea-level pressure. The first step is to compute an estimated sea-level pressure, P_E, by reducing the station pressure assuming standard atmospheric conditions: $P_E = P^* \left[\dfrac{288.}{288. - .0065 ELEV} \right]^{5.256}$. We denote the magnitude of the pressure difference DELP $= |P_E - P_0|$. Four tolerances are determined from the station elevation as follows:

$$\text{TOLA} = 0.15^* ((ELEV)^{**} 0.5) + 1.$$
$$\text{TOLB} = 0.20^* ((ELEV)^{**} 0.5) + 1.$$
$$\text{TOLC} = 0.20^* ((ELEV)^{**} 0.5) + 3.$$
$$\text{TOLD} = 0.004^* (ELEV)$$

The consistency checks can be summarized as follows:

SATISIFIES	RECEIVES
ELEV $<$ 500m, DELP $<$ TOLA	"A"
ELEV $\leq$ 1000m, $T^* \geq$ $-$ 15°c, DELP$>$TOLB	"B"
500m $<$ ELEV $\leq$ 1000m, $T^*<$25°c, DELP$<$TOLD	"D"
ELEV $\leq$ 1000m, $T^*<$$-$15°c, DELP$<$TOLC	" "
ELEV $<$ 1000m and all others	" "

2.3.4 Monitored Data

Manual quality control is handled by analysts from NMC's Meteorological Operations Division. Reports can be examined and selectively corrected, purged or retained. This includes the ability to flag a single parameter at a single level, a complete report, or a block of reports in a given area. These monitor flags are incorporated when the basic data sets of Table 1 are generated. The "purge" or "hold" flags are used in place of any existing quality marks.

3. DATA PRE-PROCESSING FOR THE OI ANALYSIS

Once the basic data sets of Table 1 are constructed, it is the purpose of the pre-processor to select the information required by the analysis and to output the data in the form expected by subsequent software, all of which is run on the CYBER 205. Preprocessing is performed on the front-end computers to facilitate unpacking of ON 29/124 data, which are in EBCDIC character form, and to minimize the volume of information to be transmitted across the data link between the front-end and the CYBER 205. The major functions of the preprocessor are (1) to read in and unpack the reports, (2) to perform rudimentary checks for time, location, completeness and quality, (3) to convert units and apply corrections and (4) to pack, block and write out the data.

3.1 Input/Output Processing

We begin by discussing the first and last functions, both of which involve input/output processing and packed formats. The input data sets have already been discussed, as has the ON 29/124 format of the input data. Reports are read in, unpacked and dealt with one at a time. Observations are processed into a fixed-length block containing 400 reports, with each report occupying 56 locations of 2 bytes each. All numerical parameters are stored as signed integer values (IBM FORTRAN: INTEGER*2). The 56 locations are partitioned

according to Table 5. The first 8 locations contain the report identification data, followed by 12 levels of 4 values each. At present, there are two "types" of reports - one for mass and moisture data and the other for wind data. Therefore, each observation report is packed into a mass and/or a wind report, depending on the information it contains.

Table 5. Report formats for OI analysis

Value	Contents	Units x Packing	Range
1	Latitude	(degrees +90)*10	0-1800
2	Longitude (positive E)	degrees * 10	0-3600
3,4,5	Report name	up to 6 alphanumeric characters	
6	Last level with data	-	1-12
7	Observation time	hours * 100 GMT	
8	OI Report type	(see table 7)	

-Mass/Moisture Report[1] -

Value	Contents	Units x Packing	Range
9,13,...53	Relative humidity (GOI)	% * 10	1-1000
	or Specific humidity (ROI)	g/kg * 10^3	
10,14,...54	Pressure (plus quality mark)	mb * 10+IQ	
11,15,...55	Virtual temperature	°C * 10	
12,16,...56	Height-standard height	m * 10	

-Wind Report-

Value	Contents	Units x Packing	Range
9,13,...53	Missing	-	32767
10,14,...54	Pressure (plus quality mark)	mb * 10 + IQ	
11,15,...55	Zonal wind component	ms^{-1} * 10	
12,16,...56	Meridional wind component	ms^{-1} * 10	

[1] For the GOI, mass/moisture reports are always in mandatory-level order, starting with 1000 mb in level 1 and ending with 50 mb in level 12, with missing levels indicated where necessary.

There is no order or structure to the data at this point. Data are processed until 400 reports have been accumulated in the block; then the block is written out. For the purpose of blocking the reports, if an observation has fewer than 12 levels, the remaining levels are coded as missing. The actual number of data levels will always be provided in the 6th position of the report identification.

Note from Table 5 that most values are stored to the nearest tenth. However, pressure level of the observation is truncated to the nearest mb, which is not crucial to the GOI, since it analyzes on isobaric surfaces. The tenths digit of the pressure is occupied, instead, by an integer quality mark. Note that there is allowance for only one quality mark per level. Therefore, it is impossible to supply separate quality marks for moisture, temperature and height. The OI integer quality marks are listed in Table 6 and the OI report types in Table 7a and 7b.

Table 6. OI Integer Quality Codes and ON 29/124 Equivalent

OI Code	ON 29/124	Meaning
0	"H"	Monitor keep
1	"A"	Correct, passed checks
2	blank	Probably correct, not checked
3	"D"	Suspect, passed with loose limits
9	"P","B","F","C"	Purged or failed checks

Table 7. OI Report Type Codes

A.   -Mass/Moisture Reports-

Code	Description
110	Upper air bogus
120	Radiosondes
130	Dropsondes reconnaissance aircraft
140	Climatology (Currently Not Used)
150	Satellite moisture bogus
161 (171)	Clear retrievals, satellite 1 (2)
162 (172)	Partly cloudy retrievals, satellite 1 (2)
163 (173)	Cloudy retrievals, satellite 1 (2)
180	Surface ships and buoys
181	Surface land reports (ROI only)
190	Surface bogus reports

B.   -Wind Reports-

Code	Description
220	Rawinsondes
221	Pilot balloon winds
230	Aircraft winds (AIREP/ACARS)
231	ASDAR (Aircraft to Satellite Data Relay) winds
232	Dropwindsondes
240 (250)	Low-(high) level cloud drift winds (US satellites)
241 (251)	Low-(high) level cloud drift winds (Indian satellite)
242 (252)	Low-(high) level cloud drift winds (Japanese satellite)
243 (253)	Low-(high) level cloud drift winds (European satellite)
270	Constant level balloon winds
280	Surface ship winds

3.2   Treatment of Remote Temperature Soundings

Profiles of temperature retrievals from polar orbiting satellites (types
161-163 and 171-173) are obtained from on-line data sets on the front-end which
are updated continuously by the National Environmental Satellite and Data
Information Service (NESDIS). Profiles of layer mean virtual temperature are
first converted to profiles of geopotential thickness. The lower level used
in each thickness computation is the 1000 mb level, where the geopotential
height is set to zero. The upper levels are the mandatory pressure levels
from 850 through 50 mb. The thickness values are stored as if they were normal
height values reported at the pressure of the upper level. No moisture retrieval
information is used at this time.

Prior to conversion and inclusion in the output data set, retrievals
are checked against several criteria. Observations must be within +/-3 hours
of the analysis time and must have a zero elevation; thus only oceanic retrievals
are used. In addition, cloudy retrievals, which use microwave channels only,
are not used in the tropics between 20°N and 20°S latitude. There are two
sources of quality information which are examined to see if a retrieval should
be excluded. The first is an internal marker (provided by NESDIS with most
retrievals) which indicates whether the retrieval has been checked and, if so,
whether it should be kept or tossed. The second is a partial set of hold/purge
flags set by the monitor during manual quality control. If either of these sources
provide a purge flag, the report is excluded. If the manual flag indicates a
hold, then the retrieval's OI quality mark is set to a 1; otherwise, it defaults
to a 2.

3.3   Moisture Bogus Profiles

NESDIS also provides NMC with a set of moisture bogus data twice a day (0000 GMT and 1200 GMT). From 400-500 oceanic points are provided covering the eastern Pacific, Gulf of Mexico, and western Atlantic at a resolution of about 250 km. At each point, a code figure (1-13 in Category 8) is provided which corresponds to a characteristic profile of relative humidity, RH. The preprocessor converts this code into values of RH at 6 pressure levels. The data values prescribed are listed in Table 8.

Table 8.   Code - Relative humidity (%) equivalents for moisture bogus.

Code (mb)	1	2	3	4	5	6	7	8	9	10	11	12	13
300	75	55	53	34	10	10	10	10	70	10	10	70	70
400	85	74	67	44	50	10	10	10	70	10	20	70	70
500	90	83	67	54	85	20	25	15	25	10	85	32	85
700	95	87	76	62	50	79	50	25	15	10	90	25	25
850	98	93	86	70	39	92	89	65	60	35	92	89	50
1000	95	90	80	72	68	84	76	73	69	45	90	76	68

3.4   Preliminary Data Checking

Data checking and quality control are not a major function of the data pre-processor; however, some validation and checking are performed. For example, if any of the following are encountered for any type of report, the report is skipped and not included in the output data file:

o latitude missing or out of range -90° to +90°

o longitude missing or out of range 0° to 360°

o observation time missing

o observation time more than 3 hours off-time

o observation type missing.

Individual levels or specific parameters are excluded or coded as missing if:

- o the quality mark indicates a "bad" value

- o the quality mark indicates a monitor purge

- o the relevant level Z, P or Za is missing

- o for Tdd, if the temperature is missing or "bad"

- o for wind, if either DD or FF is missing or "bad".

All surface land reports are excluded in the GOI. This is required to reduce the volume of data and running time of the system, but would not be required if a "super-ob" technique could be incorporated for the dense surface land network, as is being developed for the ROI.

Some data types can be excluded en mass by setting certain external input switches. These include upper-air bogus (not used in the GDAS or the ROI) and satellite temperature soundings. Satellite soundings can be omitted by specifying satellite number, location, pressure level and/or retrieval type. It is via this mechanism that retrievals over land and tropical microwave retrievals are excluded.

3.5   Units Conversion and Other Adjustments

As can be seen by comparing Tables 2 and 5, some of the data in each report must be converted into the proper form or units for the OI analysis. The latitude is converted from the range -90° to +90° (positive north) to the range 0-180. The longitude is converted from degrees west to degrees east of the Greenwich Meridian in the range 0-360°. Winds are converted from direction and speed in knots to zonal and meridional components in ms^{-1}. Dew-point depressions are converted to relative humidity (GOI) or specific humidity (ROI), but only up to 250 mb and only if the dew-point temperature is greater

than or equal to 215°K. Temperatures are converted, to virtual temperatures using the moisture information, although the GOI does not use them.

The following expressions are used for moisture-related conversions: vapor pressure:

$$ES=6.1078*EXP((17.269*T)/(T+237.3)), \text{ where } T=\text{temperature °C}$$

$$E=6.1078*EXP((17.269*TD)/(TD+237.3)), \text{ where } TD=T-Tdd=\text{dew point}$$

specific humidity

$$qs=0.622*ES/(P-.378*ES), \text{ where } P=\text{pressure in mb}$$

$$q=0.622*E/(P-.378*E)$$

relative humidity

$$RH=q/qs*100$$

virtual temperature

$$Tv=T*(1.0+0.61*q)$$

Upper-air heights are stored as "D-values" by subtracting a reference height for the pressure level in question (see Table 4). Aircraft pressure altitudes are converted to pressures using the following (FORTRAN) functions:

$$PRH(Z) = 226.3*EXP(1.576106E-4*(11000.-Z)), \text{ where } Z>11000 \text{ meters, or}$$

$$PR(Z) = 1013.5*((288.-.0065*Z)/288.)**5.256, \text{ where } Z\leq11000 \text{ meters}$$

The latter expression returns the standard atmosphere pressure for a given height and is used for the occasional low-level aircraft report. Satellite cloud-drift winds are reported in the same format as aircraft winds (category 6, see Table 2), but their valid pressure, if available, is coded in Category 8. If the pressure is not available, the pressure altitude is converted as described above. Temperature information from either aircraft or satellite cloud-drift winds is not used.

Surface wind data over the oceans are corrected for the effects of friction in an attempt to get an equivalent geostrophic value. This procedure is based on the marine boundary-layer model of Cardone (1964) and Druyan (1972). The reported direction, DD, and speed, FF, are adjusted as follows:

FFADJ = 1.91*FF-5.97

DELDD = 26.5-17.3*FF/FFADJ+.04*(ABS(YLAT)-35.), where YLAT is station latitude,

DDADJ = DD+DELDD*SIGN(1.0,YLAT)

If the new speed, FFADJ, is less than or equal to zero, the wind is replaced by a calm wind. Surface wind data from land stations are not used because a reliable conversion scheme is not available. They would be of low utility in any event, because surface wind is not an analysis variable. The special effort to get oceanic winds is justified by the lack of data over the oceans, and the fact that winds are very useful in extending the influence of height data via the multivariate nature of OI analyses.

Values of sea-level pressure are output in place of the height D-values for all types of surface data. They are packed in this location to the nearest tenth of mb, and truncated to the nearest whole mb in the normal location for pressure. These values will later be converted to 1000 mb heights. If the station pressure is reported instead of sea-level pressure, and the station elevation is less than 7.5 m, it is used as if it were the sea-level pressure. For the GOI, the default quality mark is 2 for all surface reports, except for Northern Hemisphere surface bogus, which receives a value of 1.

Rawinsonde heights and temperatures at mandatory levels from 100 mb
and above are corrected for the effects of shortwave solar radiation. At
present, only the following instrument types are corrected for their day-
light ascents: USA-external thermistor, USA military AN/AMT-4 external
thermistor, Finnish Vaisala, Japanese code-sending, East German Freiberg,
United Kingdom Kew and USSR A-22. The corrections of McInturff and Finger
(1968) are tabulated by pressure level and solar elevation angle, which is
computed from the day and time of the balloon ascent. A longwave correction
is applied to all instruments at the 10 mb level. The root-mean-square
(RMS) corrections are not large, rarely exceeding 60 m. The tabulated
values are in need of revision, since they are based on values which have
since been updated and expanded by McInturff et al. (1979).

The storage of rawinsonde data is now straight forward. Mandatory
level data for mass and moisture are extracted from Category 1 (see Table
2), with missing or rejected data stored as missing. Wind data are extracted
in a similar fashion, except that the number of missing levels are counted.
When wind data from Category 1 are exhausted, the number of levels required
to complete the report with 12 levels of data are extracted from the re-
maining categories in the following order. The tropopause level (Category
5) wind data are considered first, and if they are not already present in
the mandatory-level data, they are included. If room still remains, then
significant-level winds-by-pressure or winds-by-height, whichever are more
plentiful, are used to fill the remaining levels. The surface-level wind
is not included, however, as it is contaminated by nonrepresentative effects.
Finally, before the combination of winds is added to the output block, the
levels of data are arranged in order of decreasing pressure.

4. CALCULATION AND QUALITY CONTROL OF FORECAST ERRORS

4.1 Forecast Error Calculation

The observational data set created by the data preprocessor is transferred to the CYBER 205, where it is processed into a file named "FERR" (which stands for Forecast Error). First, the data blocks are read in and converted into CYBER 205 format. The reports are then stored in a large contiguous data buffer, with the missing levels no longer included. This is the master data buffer. Next, the first guess fields of height, temperature, relative humidity and wind are read in. These values are located at the intersections of a 2.5° latitude-longitude grid and, under normal circumstances, have been provided by a 6-hour forecast from the GDAS. Therefore, they have been post-processed from the 12-layer sigma domain of the model to the 12 mandatory-pressure surfaces.

The observations are now processed into 72 strips covering 2.5° latitude each, starting at the south pole and extending to the north pole. The data in each strip are ordered by increasing longitude from Greenwich eastward. The format for each strip is the same as that for the blocks of input data, i.e., 400 reports per strip with 56 values per report.

Observations of sea-level pressure are processed as follows. First, the temperature and height fields from the first guess at 1000 mb are interpolated horizontally to the report location. These values are used to compute a first guess sea-level pressure $P_{OG} = 1000.*EXP(g*Z_{1000}/(R*T_{1000}))$, where $g = 9.8 \text{ ms}^{-2}$ and $R=287.05 \text{m}^2\text{s}^{-2}\text{K}^{-1}$. Next, the observation P_O and first guess P_{OG} are differenced, and the pressure residual converted to a 1000 mb height residual:

$$RESZ = \frac{R*T_{1000}}{g*1013.5} * RESP, \text{ where } RESP = P_O - P_{OG}.$$

For other types of data, residuals are calculated as follows. For each value in each report of each strip, a value of the first guess at that location is generated by interpolation. The observed residual is formed by subtracting the first guess from the observed value. The observed values are then replaced by the residual values.

4.2 Gross Error Check

Prior to the analysis, the residuals are subjected to several quality control steps. The first is designed to eliminate meteorologically unreasonable reports, and is often called the "gross-error check". It compares the magnitude of each residual with a forecast-error standard deviation σ computed from a long series of cases. These have been compiled at all mandatory levels, for 5 latitude bands, and are listed in Table 9. The degree of tolerance allowed for an upper-air residual to pass the gross check depends on the OI quality mark of the observation. (These limits are given in Table 10.) Note that higher quality observations are allowed a greater tolerance than those of poorer quality. Also, observations manually flagged for retention and given a quality mark of 0 are not checked at all, and are included unconditionally. Otherwise, for the wind, each component is checked individually, and if either one fails, they are both tossed by setting the residuals to missing. For the mass observations, each data type is considered independently, and is set to missing (tossed) if it fails the check.

Those residuals which survive the gross check are then compared to another more stringent tolerance level (See Table 10) to see if they should be flagged as questionably "large" residuals. This designation is used in the internal consistency check and is effected by simply adding 4 to the existing quality mark.

TABLE 9

FORECAST ERROR STANDARD DEVIATIONS
σ-VALUES

Latitude	90S-10S			10S-10N			10N-30N			30N-50N			50N-90N			90S-90N
Pressure (mb)	z (m)	u (m sec^{-1})	v	z	u	v	z	u	v	z	u	v	z	u	v	T(°K)
1000	25.4	4.2	4.1	14.7	3.5	3.3	19.4	4.1	3.9	19.7	4.6	4.9	21.6	4.4	4.3	4.0
850	25.4	3.9	3.7	15.5	3.2	2.8	16.7	3.7	3.4	19.6	4.1	4.3	19.4	3.6	3.6	4.0
700	26.8	4.2	4.1	19.3	3.6	3.3	18.7	4.1	3.6	19.7	4.2	3.8	20.8	3.6	3.6	4.0
500	33.5	4.8	4.6	25.4	4.3	3.5	25.6	4.8	4.2	23.2	4.8	4.6	25.8	4.4	4.4	4.0
400	39.2	5.5	5.7	29.6	4.1	4.0	32.4	5.3	4.9	27.2	5.6	5.3	29.9	5.3	5.0	4.0
300	47.0	6.6	6.8	40.1	5.8	5.3	42.3	6.5	6.1	34.1	7.1	6.4	36.5	5.7	5.5	4.0
250	50.6	9.5	7.3	49.4	5.8	5.2	47.3	7.3	6.9	37.4	6.6	6.6	39.2	5.4	5.3	4.0
200	53.5	7.3	7.0	55.3	7.8	5.8	56.3	7.6	7.6	41.6	6.7	6.2	42.0	4.4	4.3	4.0
150	57.3	6.5	6.7	61.4	8.2	5.7	67.4	7.3	7.1	46.7	5.9	4.9	48.3	3.7	3.7	4.0
100	69.8	7.0	6.4	78.0	8.8	7.1	80.1	7.7	6.2	55.2	5.1	4.1	59.4	3.7	3.6	4.0
70	77.7	6.6	5.5	93.2	7.7	4.8	100.0	5.7	5.3	64.2	7.1	4.9	71.8	5.5	5.1	4.0
50	90.1	8.3	8.0	108.0	9.9	8.3	101.0	8.7	8.6	78.3	8.2	6.7	88.1	7.6	7.4	4.0

Table 10. Limits used in Gross Error Check

Data Type	Quality Mark	GOI Limits Toss if>	FLAG if>	ROI Limits TOSS if>	FLAG if>
Upper-air					
Z, T, U, V	1	7 σ	3 σ	5.5 σ	2.5 σ
	2	5 σ	3 σ	4.5 σ	2.5 σ
	3	4 σ	3 σ	3.5 σ	2.5 σ
Surface					
U, V	1	5 σ	2 σ	3.0 σ	1.5 σ
Surface					
Z, T	2	4 σ	2 σ	3.0 σ	1.5 σ

Prior to writing out each strip of residuals, each report is checked for missing data. If relevant information at all reported levels has been tossed and/or is missing, the report is skipped. Checks are also made at this time for duplicate reports. If the report identification (the first 8 values of Table 5) for two reports is the same, the second occurrence of the report is skipped as a duplicate. Once this thinning is completed, the strip of residuals is written out for later use in the analysis.

5.   HORIZONTAL CONSISTENCY CHECK

Prior to actually performing any analyses, the residuals are checked for horizontal consistency. This is achieved by checking each residual against its neighbors, hence the term "buddy check". The check is univariate, in that values are checked only against neighbors of like type (z, u or v); and it is more or less two-dimensional, since the residuals are grouped into vertical layers by pressure.

First, all of the strips of residuals are read sequentially into a large data array. The starting addresses of each 2.5° strip, and of each residual are also saved. Next, we define another latitude-longitude grid which is used for buddy-checking the data. It is an "equal-area" grid, where points are evenly spaced in latitude every 5°, but the separation in longitude varies with latitude such that the eastwest spacing in physical distance is conserved. Thus, the number of points around a latitude circle decreases as one moves from the equator to the poles. This grid is used to define reference points about which data will be collected and checked.

At each reference point, all residuals within 7.5° latitude and within an equivalent distance in longitude are collected. Thus, data from six strips, three on either side of the reference point, will be selected; data from each strip will span at least 15° (~1558 km). Since the reference grid points are 5° apart and the collection radius is 7.5°, all data on the globe will be considered at least once. We also define an internal area with a smaller radius, which covers the whole globe but minimizes the areas of overlap (see Fig. 1, and later discussion).

All the data collected are vertically sorted into groups of common data type (Z, u, or v) for each of the 12 mandatory-pressure levels. Off-level data such as significant-level winds or aircraft winds are included with like data at the nearest pressure level.

Each group with at least 3 residuals is subjected to the checking pro-cedure discussed below. If there are only one or two residuals in a group, a check is made for a questionably large residual. Recall that those were flagged as part of the gross-error check, and are indicated by the quality mark. If either one or both of the residuals is flagged as such, the suspect value is removed from further consideration by either the buddy check or the analysis.
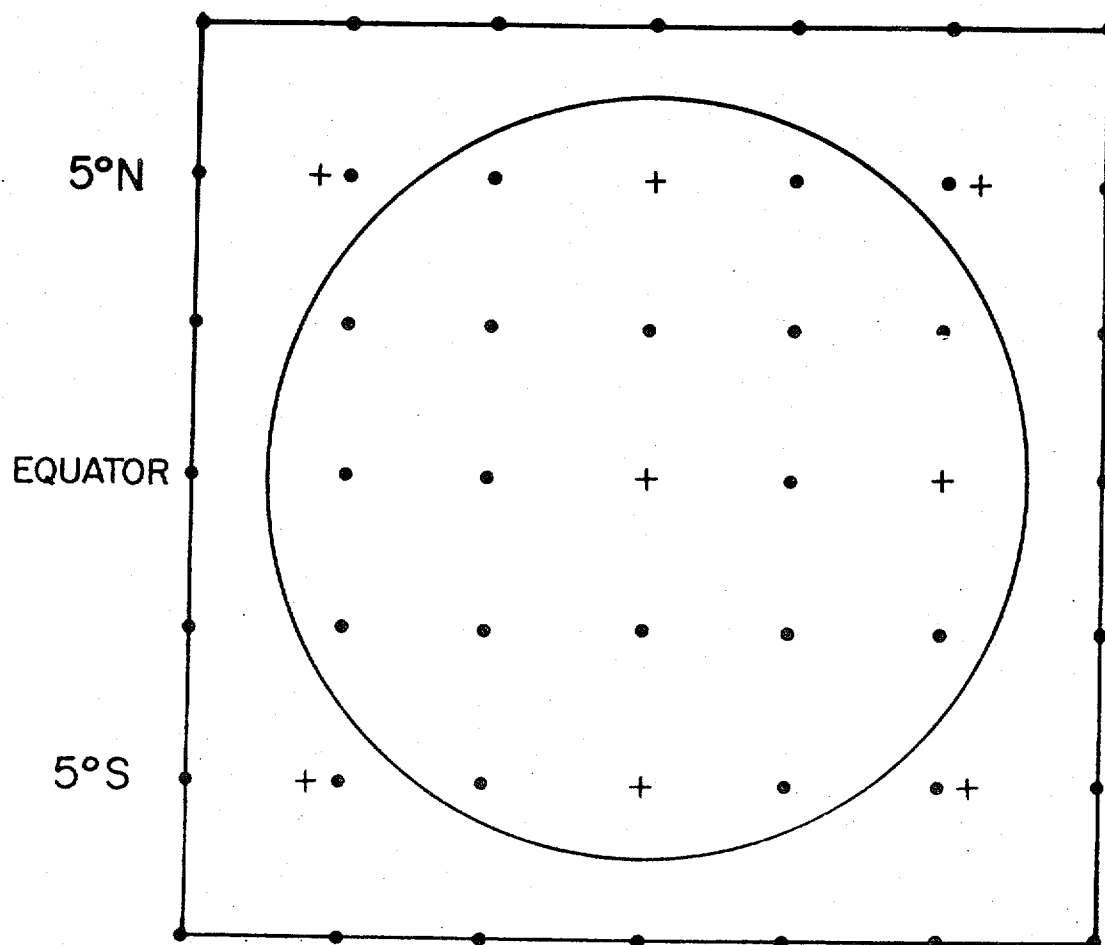
Fig. 1. Relationship of 2½° storage grid (points), buddy check 5° reference grid (+), internal area for eliminating data (circle) and collection area for data influencing the screening (square).
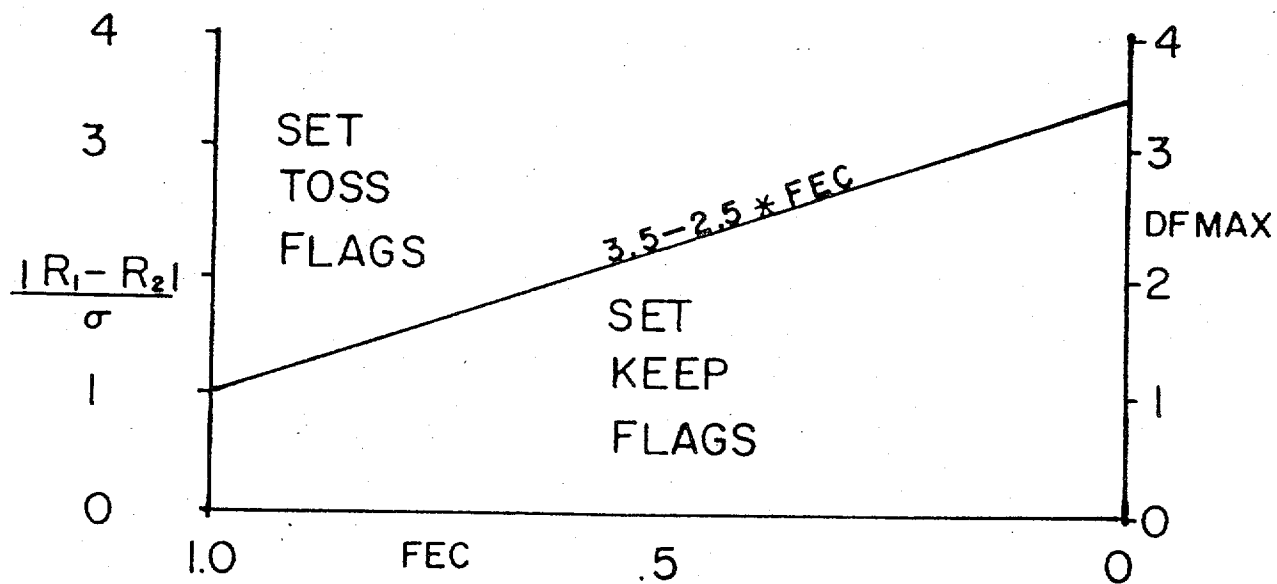

Fig. 2. Schematic representation of screening criteria (see text).

For groups with three or more residuals, the forecast-error correlation (FEC) is computed as a function of separation distance for each pair of values in the group. The horizontal function for the ZZ correlation has the Gaussian form FEC=EXP(-Kd^2), where d is the separation distance in km and K=2.E^{-6} km^{-2}. The autocorrelation functions for wind (UU and VV) are computed geostrophically from the height-height (ZZ) correlation (Bergman, 1979). Next, each pair of residuals is compared and the magnitude of the difference is normalized by the appropriate standard deviation of the forecast error from Table 9. This This normalized difference, $\frac{|R_1 - R_2|}{\sigma}$ in Fig. 2, is compared to a limiting value, DFMAX=(3.5-2.5*FEC). If the difference exceeds the limiting value, then "purge" flags are set; otherwise, "hold" flags are set. If the difference is equal to the limiting value, no flags are set.

Fig. 2 is a schematic representation of this process. When two residuals are far apart, their correlation FEC is small, but the limiting value DFMAX is large. When two residuals are close together, on the other hand, FEC is large, DFMAX is small, and the residuals are expected to agree, at least to within one standard deviation.

The setting of flags depends on the quality marks of the residuals. Recall from Table 6 that lower values imply better quality. If the quality marks are the same, then both residuals receive either a purge flag or hold flag. For setting purge flags, the residual with the larger quality mark (the more questionable ob) receives the purge flag and the residual with the smaller quality mark receives no flag. Correspondingly, when setting hold flags, the residual with the smaller quality mark (the better ob) receives the hold flag. All hold flags and purge flags are stored separately, corresponding to the appropriate pair of residuals.

The total number of hold and purge flags accumulated by a residual determines whether or not the report is eliminated from further consideration. Any value which receives 2 or more hold flags is automatically retained. Any residual which has been flagged as questionably large and does not have at least 2 hold flags is removed. The final step is to eliminate residuals with two or more purge flags by setting the residual to missing. Only residuals within the internal area, defined by the 0.8 value of the ZZ forecast error correlation, are actually rejected. Although data from the entire collection area contribute to and receive flags, the data outside of the internal area are not checked at this reference point for elimination. All data fall within the internal area of at least one reference point.

Elimination of data proceeds in the following manner. The internal group of residuals is checked for the value with the largest number of purge flags. If this total is two or more, this residual is set to missing and all flags which this residual set on other data are removed. Thus, the removal of an observation will generally result in fewer flags on the remaining observations in the group. Next, the remaining observations are searched again for the one with the greatest number of purge flags. The process is repeated until there are no residuals remaining with two or more purge flags. One should note, in conclusion, that if a wind component is eliminated, its other component is as well. If, of course, the v-component were rejected, it would be impossible at this point to eliminate any negative influence its corresponding u-component might have had. It is also clear that this entire process will not handle the case where there are three or more rogue residuals which corroborate one another. Each would have at least two hold flags and none would be eliminated.

The general philosophy of the buddy check, we feel, is rather liberal in that only two corroborations are required to retain an observation. On the other hand, isolated observations with large residuals are dealt with most conservatively. They are simply eliminated. This demonstrates our lack of knowledge on how to adequately incorporate these outliers when they deviate so much from the first guess field. Our rationale for deleting such observations is based on the knowledge that such observations can cause severe problems in the forecast.

In practice for the GOI, the buddy check is done in two separate steps. During the first pass through the data, all remote sounding data are excluded from consideration, since they are, at this point, anchored to the first guess 1000 mb height. Only the wind data and the 1000 mb height data are checked at this time. Then the 1000 mb analyses of height and wind are performed, using the buddy-checked data. Once the height corrections to the first guess field are available from the analyzed 1000 mb height field, it is possible to correct the remote sounding residuals. The analyzed correction is first interpolated to the location of each remote sounding, and is then added to each height residual in the profile. At this point, all of the 850 mb to 50 mb height data are buddy checked. The remainder of the upper levels are then analyzed using the complete data base.

The average number of residuals rejected in both the gross error check and the buddy check for a 25-day period in the spring of 1983 is given in Table 11.

Table 11. Average number of residuals tossed by Gross Error check and buddy-check. May 27, 1983 through June 20, 1983.

DATA TYPE	0000 GMT			0600 GMT		
	TOTAL* REPORTS	GROSS CHECK	BUDDY CHECK	TOTAL* REPORTS	GROSS CHECK	BUDDY CHECK
MASS DATA						
Rawinsondes	837	6	46	51	0	0
NOAA-6	–	–	7	–	–	4
Clear	177	0	–	121	0	–
Partly Cloudy	311	4	–	155	0	–
Cloudy	258	16	–	146	0	–
NOAA-7	–	–	19	–	–	12
Clear	290	2	–	214	5	–
Partly Cloudy	481	10	–	373	0	–
Cloudy	329	18	–	344	21	–
Ships	827	24	12	689	20	8
BOGUS						
Pressure	466	6	22	205	2	3
Moisture	418	0	0	0	0	0
WIND DATA						
Rawinsonde	873	23	160	56	2	10
Aircraft						
Standard	702	8	20	673	10	14
ASDAR	21	0	0	42	0	0
Satellite winds						
Low-level	631	2	–	0	0	0
High-level	561	5	–	0	0	0
Ships	757	17	54	645	12	45

* A sounding (up to 12 levels) is considered a single report in this total.

6. DIFFERENCES BETWEEN GOI AND ROI DATA HANDLING

The new ROI analysis is very similar in some ways to the GOI, especially in terms of the fundamental mechanics of an OI analysis. The primary difference is in the domain of the analysis and the data used. The GOI is global; analyses of z, u and v are generated on an equal-area Kurihara-type grid on 12 mandatory levels, with only a few selected significant-level winds used. The ROI is hemispheric; analyses of z, u and v are generated on a 2° longitude by 1.5° latitude grid which is thinned in a way that is not important to this discussion. The vertical structure is defined in terms of the sigma structure of the nested-grid model (NGM), which has 16 layers - the first twelve of which are below 250 mb and require moisture analyses. Since this stratification is much finer in the lower troposphere than the GOI, significant-level mass, moisture and wind data are used throughout.

The generation of the input data sets of Table 1 is identical for the GOI and the ROI. The only difference is the actual time of day when the analysis suites are run. The ROI runs off the data available at HH+2:15, the operational GOI at about HH+3:40, and the GDAS at about HH+8:00, where HH is the main synoptic time of 00 or 12 GMT. Since there will be less data received at the earlier data cut-off time, there is less data available to the ROI than the GOI. Receipt times for the critical areas of the ROI were examined with respect to the availability of significant level data; about 70-80% have usually arrived by HH+2:15.

The data preprocessor for the ROI has the additional requirement of processing all of the significant level data from rawinsonde reports. All upper-level data (categories 1-5, Table 2), are processed as follows. The mandatory-level data through 10 mb are extracted and merged with the significant

level temperature data through 100 mb (category 2) and the tropopause data (category 5). Geopotential heights are generated at the significant levels by hydrostatic integration between the reported mandatory levels. The integration uses the virtual temperatures determined from the dew-point depression data. The OI quality mark (Table 6) for the computed height is set according to the greater of the three values used in the calculation - the height and temperature below and the temperature at the level in question. In addition, there is an integration performed for each layer with a valid mandatory-level height above the last significant level. This integrated height is compared to the reported height at the mandatory level. If the magnitude of the difference is greater than $(3.5-P/50.)$, where P is the mandatory level pressure in mb, then a problem is assumed to exist with the reported temperatures in this layer. In the case of such a failure, all computed heights and significant-level temperatures, down to the last mandatory level are given a quality mark of 9. This will prevent them from being considered any further.

The next step is to merge the winds-by-pressure (category 3) with the existing profile. Duplicate pressure levels are deleted, provided all data are complete. Temperatures are generated at these wind levels by linear inter-polation with respect to the logarithm of pressure, and heights are computed by integration. Similarly, winds-by-height (category 4) are merged with the existing profile, in their proper place, by reported height. For wind profiles without category 1 or 2 mass data (PIBAL winds), a standard atmosphere variation of height, temperature and pressure is assumed. These winds receive quality marks of 3 and their standard atmosphere values of height and temperature are flagged with 9's to ensure they will not be considered by the analysis. For complete profiles with reported mass data, a pressure is generated at the reported height level by assuming a quadratic variation of height with respect to the logarithm of pressure. Finally, temperatures are interpolated as before.

At this point, every level in the fully merged profile has a height, temperature and pressure. The final step is to complete the moisture and wind information at each level. Those levels with missing dew points or winds have values computed by linear interpolation with respect to the logarithm of pressure, provided there are bracketing levels with reported information. Several inter- polated levels can be generated from a single pair of reported values. As before, the OI quality mark is set to the largest value (most questionable) of those being used to generate the interpolated value.

In practice, this procedure of merging the various categories of rawin- sonde information is also used to generate a complete set of values at regular pressure levels every 25 mb. This is achieved by adding the desired pressure levels to the list of winds-by-pressure. Processing of temperatures and heights is performed as described above, even though there are no winds at these added levels. All the information concerning the temperature structure is available from the category 1 and 2 data, so the resulting values are truly representative. Winds and moisture data are generated in the final step, again, when all reported data have been processed.

The merged sounding, with all levels complete, is packed into the format of Table 5. Since the format allows only 12 levels of information, the full profile is written out in segments, with up to 12 levels per segment. Each segment has the total number of levels coded in the sixth parameter of the report identification. This allows merged profiles to be uniquely identifiable. Each successive segment has unity added to its time of report in the seventh parameter. This keeps the separate pieces from being tossed as duplicates. When reports are grouped into blocks, a check is made to insure that the segments of a merged report, both mass and wind, will fit into the current block and

not spill over into the next block. If spillover would occur, the current block is simply written out with fewer than the maximum of 400 reports and a new block is started with the merged report.

When the data blocks are processed on the CYBER 205 (by the FERR file generating code), a special check is made for the merged profiles. The separate pieces of the profile report are re-merged and the data to be used by the ROI is extracted. The ROI can use data at any location and pressure - unlike the GOI which assumes the data are in mandatory-level order. Therefore, all levels of the complete merged profile could be used. However, the distribution and number of levels available varies greatly from one report to another. This variation makes it difficult to predict exactly what data might be picked during the data selection procedure. Also, the large volume of data requires more time to process. Therefore, the ROI uses only the data profile that was generated in the preprocessor on fixed pressure levels every 25 mb. This amounts to up to 36 pieces of information (up to 20 mb) which are extracted from the reconstructed profile. These are made available in three segments of 12 levels each for both the mass and wind reports from a single rawinsonde.

The first guess used by the ROI is the same 6-hour forecast from the GDAS, except that it is not on pressure surfaces or the 2.5° latitude-longitude grid. Instead, the 12-level forecast in spectral coefficient form has been converted to a 16-level representation on the full 2° longitude by 1.5° latitude ROI analysis grid. During this conversion, the terrain for the high resolution NGM is incorporated into the first-guess fields through an adjustment of the surface pressure. Residuals are computed as for the GOI, except that vertical interpolation is required for nearly all of the data. Temperature, moisture and wind are interpolated linearly with respect to the logarithm of pressure,

while heights assume a quadratic variation. Limits used in the gross error check are listed for the ROI in Table 10.

Whereas the GOI utilizes only surface reports of sea-level pressure, converted to 1000 mb height residuals, the ROI analysis uses more of the surface report and in a different manner. Backtracking to the data preprocessor, the reported surface temperature, moisture and elevation (considered the height observation) are included (1) if the station pressure is reported from a land station or (2) if the sea-level pressure is reported from ships or from stations whose elevations are less than 7.5m. Land-station reports of sea-level pressure are also included in the manner described for the GOI. Surface land reports are required to be within 1.5 hours of the analysis time versus 3 hours for ships. This thinning measure will be deleted when the super-ob techniques is implemented. When extrapolation of surface data located below the first-guess surface pressure is required, the toss-out limit for the gross-error check is decreased by one standard deviation.

The first step in the ROI is to perform a univariate analysis of the 1000 mb height. This analysis uses primarily the sea-level pressure reports in the form of 1000 mb height residuals. Once the 1000 mb analysis is complete, values of sea-level pressure are computed. The heights are used to adjust the satellite sounding residuals to the proper 1000 mb level. The upper-air analyses use the surface elevation and pressure in the form of a geopotential residual. In fact, all surface data are used directly at their reported pressure, since the OI in sigma will distribute their effect in the vertical.

7.   PLANS FOR THE FUTURE

As a result of the effort to compile the information for this paper, we decided to review the various aspects of quality control being performed on the data base prior to and during the OI data pre-processing.

We are particularly interested in exploring some ideas of L. S. Gandin which, according to recent personal communiction with Gandin, have been used successfully in the Soviet Union. The basic concept involves parallel, rather than sequential, testing: each datum is subjected to a battery of tests, and only after all are completed is a decision made to retain, reject, or correct the datum.

A revised pre-analysis quality control procedure is being considered for the GOI along the lines of the procedure of Lorenc (1981) in use at ECMWF. It is expected to be tailored to the future CYBER 205 configuration of the GOI, in which multiple matrices will be solved simultaneously.

A new routine to correct radiosondes for solar radiation effects should be tested soon. It is based on the tabulated height and temperature corrections for nearly all currently used instruments contained in McInturff, et.al.(1979). The surface ship wind adjustment is being examined with respect to low wind speed performance. A simple algorithm to construct "super-obs" of nearly coincident surface and aircraft data is being developed. The possibility of a continuous updating of the forecast error standard deviation tables is also being considered.

8. REFERENCES

Bergman, K., 1978: Role of observational errors in optimum interpolation analysis. Bull. Am. Meteor. Soc., 59, 1603-1611.

_____ , 1979: A multivariate optimum interpolation analysis system of temperature and wind fields. Mon. Wea. Rev., 107, 1423-1444.

Cardone, V., 1969: Specification of the wind distribution in the marine boundary layer for wave forecasting. GSL Rep. TR-69-1. Dept. of Meteorology and Oceanography, New York University.

Dey, C. H., and L. L. Morone, 1985: Evolution of the NMC Global Assimilation System: January 1982-December 1983. Mon. Wea. Rev, 113, 304-318.

Druyan, L. M., 1972: Objective analysis of sea-level winds and pressures derived from simulated observations of a satellite radar radiometer and actual conventional data. J. Appl. Meteor., 11, 413-428.

Kistler, R. E. and D. F. Parrish, 1982: Evolution of the NMC data assimilation system: September 1978-January 1982. Mon. Wea. Rev., 110, 1335-1346.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon. Wea. Rev., 109, 701-721.

McInturff, R. M. and F. G. Finger, 1968: The compatibility of radiosonde data at stratospheric levels over the Northern Hemisphere. Wea. Bureau Tech. Memo. WBTM DATAC 2, pp. 61.

McInturff, R. M., F. G. Finger, K. W. Johnson and J. D. Laver, 1979: Day-night differences in radiosonde observations of the stratosphere and troposphere. NOAA Tech. Memo. NWS NMC 63, pp. 47.

McPherson, R. D., K. H. Bergman, R. E. Kistler, G. E. Rasch and D. S. Gordon, 1979: The NMC operational global data assimilation system. Mon. Wea. Rev., 107, 1445-1461.

9. ACKNOWLEDGMENTS